Underwater Pose Estimation Relative to Planar Hull Surface Using Stereo Vision

Dongha Chung[1], Seonghun Hong[2], Jinwhan Kim[1,2]
[1]Department of Mechanical Engineering
[2]Robotics Program
KAIST, Daejeon, 305-732, Korea
(Tel : +82-42-350-1579; E-mail: (chungdongha, sh.hong, jinwhan)@kaist.ac.kr)

***Abstract*—For creating a precise visual map by autonomous ship-hull inspection using an unmanned underwater vehicle, it is a crucial capability for the vehicle (or camera) to maintain a pose relative to the hull surface. In this study, a relative pose estimation algorithm is introduced using a stereo vision system. The proposed approach utilizes 3D point cloud data that can be generated by a sparse feature matching technique between a pair of stereo images. The relative pose information can be obtained by applying a surface normal estimation algorithm for the 3D points. Experimental results using underwater images is shown to verify the practical feasibility of the proposed approach.**

***Keywords*—Unmanned underwater vehicle, stereo vision, autonomous hull inspection**

1. Introduction

Periodic ship-hull inspection is an important task to check the damage and biofouling of the hull surface in order to prevent some accidents and degradations of ship performance. Conventional hull inspection performed by human divers possesses apparent limitations in terms of the efficiency and safety of the task due to the harsh underwater environment. Although remotely-operated vehicles have been used to assist and/or replace the divers' work, skilled operators and their intensive work are needed to obtain reliable inspection results.

With recent advances in technologies for vehicle autonomy, the use of autonomous underwater vehicles can be a useful means for the ship-hull inspection mission. The operational safety and practical performance of the task can be greatly improved by performing the autonomous navigation and automatic visual mapping [1,2]. In particular, it is highly desirable for the vehicle (or camera) to achieve the capability to maintain a pose relative to the hull surface in order to perform the precise visual mapping (i.e., visual mosaicking).

In [3], a forward-looking Doppler velocity log (DVL) is mounted on the same servo tray of the camera module, and its measurements are utilized in order to remain orthogonal to the ship-hull surface. DVL can provide acoustic measurements useful for identifying local planes, but the ship-hull images must be collected at very close range, and it can make the DVL measurement inaccurate. Moreover, the real-time tilt control of the DVL can cause additional difficulties in system design and implementation (e.g., increasing the size and required capacity of the tile actuator).

In this study, a stereo vision system is suggested to estimate the vehicle's pose relative to planar hull surface. The proposed algorithm utilizes 3D point information obtained by applying a series of stereo matching techniques, and the relative pose of the vehicle is estimated by introducing a surface normal model using the 3D points. Experimental results demonstrate the practical feasibility of the proposed approach.

The rest of this paper is organized as follows. Section 2 provides the stereo matching algorithm, and Section 3 describes the detailed procedures of the normal vector estimation. In Section 4, experimental test and the results using underwater images are shown and conclusions are presented in section 5.

2. Stereo Correspondence

To find the stereo correspondence and obtain 3D point information from it, block matching approaches have been commonly used whose measures are defined based on pixel intensity such as the sum of absolute or squared differences. However, matching underwater images is very challenging, mainly because of the monotonic or repetitive pixel values due to the large attenuation and scattering of visible light in water, and the performance of the conventional methods is often not be satisfactory. Hence, we employ a sparse feature extraction and matching algorithm considering the epipolar geometry between a pair of stereo images.

2.1. Feature Extraction with SURF

A variety of algorithms to detect salient and sparse features from a set of images have been proposed, such as the Harris corner detection [4], scale invariant feature transform (SIFT) [5], speeded up robust features (SURF) [6]. In particular, the SIFT feature has been widely applied in various applications in the computer vision community because of its high detection performance and the scale and rotation invariant properties of the detected features. However, the high computational burden of SIFT may not be acceptable for the real-time pose estimation framework. For this reason, the SURF algorithm is employed for our application, which is also commonly used for underwater image matching. In addition, to achieve a higher computational speed, we simplify the scale space of SURF by using a single octave, considering that the scales of the stereo images acquired from a planar hull surface remains almost the same.

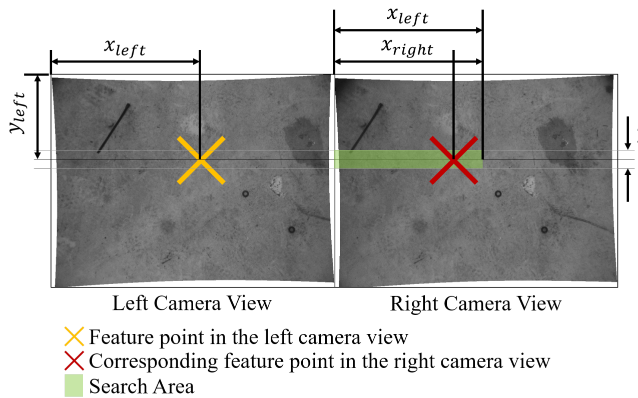
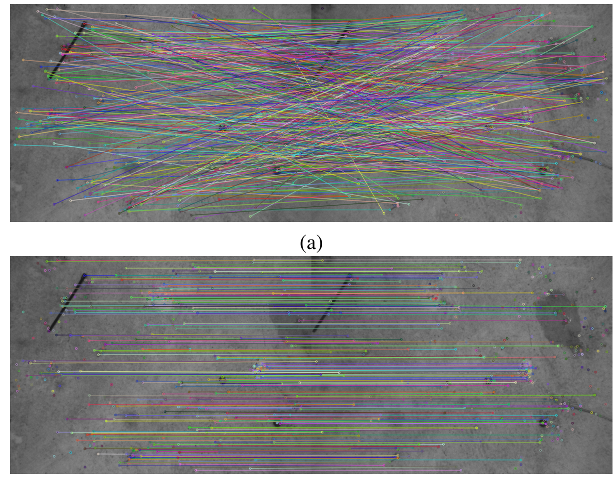


Fig. 1: A illustration of the proposed stereo matching approach



(a)

(b)

Fig. 2: A comparison between the naive stereo matching method (a) and the proposed method (b)

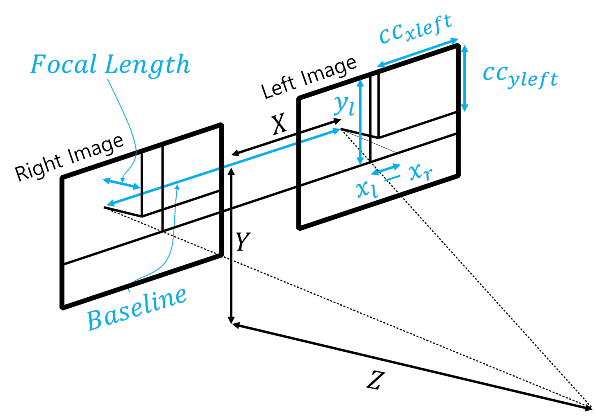


Fig. 3: Geometrical relationship between the world coordinate of a point and corresponding points in the images.

2.2. *Feature Matching*

Stereo matching is then performed using the extracted SURF features. Figure 1 depicts the proposed stereo matching method. By applying a stereo calibration algorithm, we can utilize the fact that the matched point of a feature point in one image lies on the same horizontal line in the other image by the epipolar constraint. Here, a search area is established considering the existence of errors in the camera calibration method and the uncertainty of extracted feature location. The correspondence for each feature point is determined, satisfying the following constraints:

$$\begin{aligned} &x_{left} > x_{right} \\ &|y_{right} - y_{left}| < \epsilon. \end{aligned} \tag{1}$$

Here, x_{left} and y_{left} are position coordinates of the feautre in the left image, and similarly, x_{right} and y_{left} are the feature position in the right image. By reducing the search area as shown in Fig. 1, the computation speed in the feature matching step can be substantially improved. In addition, Fig. 2 shows a comparison between the naive stereo matching and the proposed matching approach. From the results, we can observe that the proposed approach is advantageous in reducing the mismatched correspondences.

After finding the correspondence between a pair of stereo images, the disparity is calculated by subtracting the horizontal position of the corresponding point in the right image from that in the left image. The disparity map is then converted into 3D point cloud data. As shown in Fig. 3, the 3D position of each feature point can be calculated using the geometric relationship between the image coordinate and global coordinate with the pinhole camera model. The 3D position of each correspondence point is calculated as follows:

$$X = \frac{B}{x_{left} - x_{right}}(x_{left} - cc_{xleft}) \tag{2}$$

$$Y = \frac{B}{x_{left} - x_{right}}(y_{left} - cc_{yleft}) \tag{3}$$

$$Z = \frac{B}{x_{left} - x_{right}}F \tag{4}$$

where B represents the baseline length of the stereo camera, and F indicates the focal length. cc_{left}, cc_{left} are the focal center coordinate of left camera. By repeating this process for all the matched features, we can obtain the 3D point cloud data of the current scene.

3. Relative Pose Estimation

Although the line search method for matching SURF features outperforms global search method, there still exists mismatched features as shown in Fig. 4. These mismatched results provide wrong position estimate of the feature point in 3D coordinates, since the position is the function of disparity shown in Eq. 4. Since these mismatched points may affect the estimation of the plane, RANSAC with the plane model is applied to eliminate such mismatched points. The RANSAC algorithm with the plane model finds the inlier points whose shortest distance to the plane is smaller than a value. Figure 5 shows the result of applying RANSAC with the extracted 3D point cloud. After rejecting the outliers, SVD is applied to the

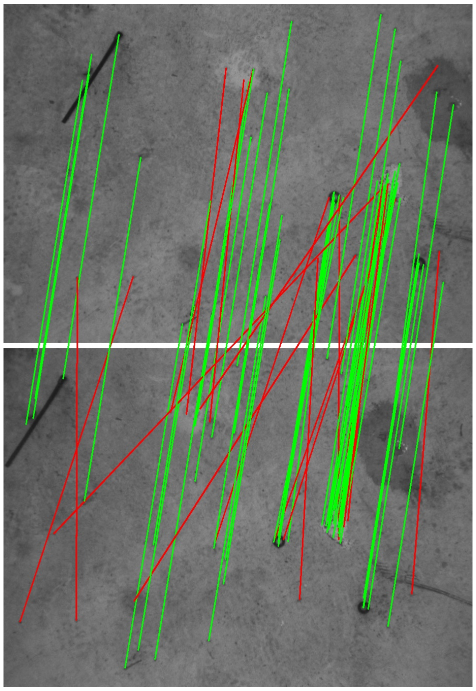


Fig. 4: Matching results : inlier pairs in green lines and outlier pairs in red lines

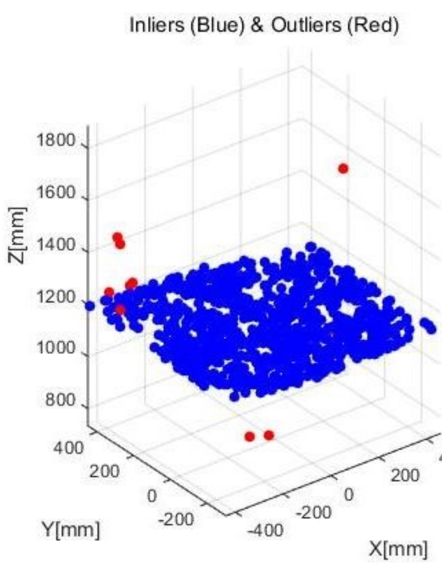


Fig. 5: Result of RANSAC with the plane model. The inliers are shown in blue, and outliers are shown in red.

remaining inliers to get the plane equation where the square sum of the shortest distance to the plane of each points is minimum. With this plane equation, we can decide the yaw and pitch angles of the vehicle and the camera tilt unit, and also the distance from the plane. Using Eq. 5, the distance from the plane, and the yaw and pitch angles to the plane can be measured as Eq. 6.

$$ax + by + cx + d = 0 \tag{5}$$

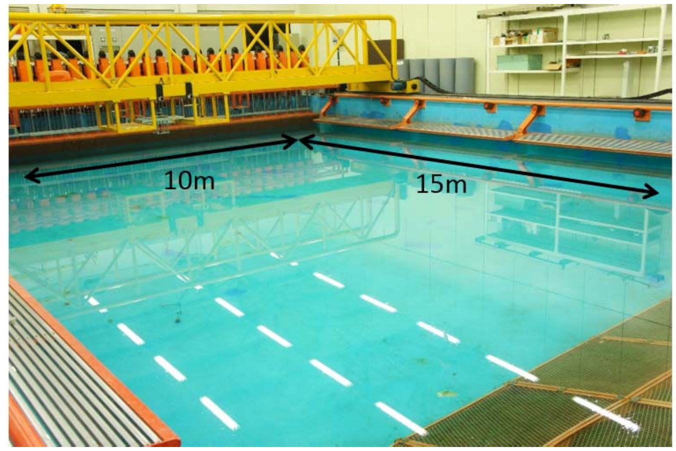


Fig. 6: Indoor water tank located in KAIST.

Fig. 7: Stereo camera unit for experimental validation.

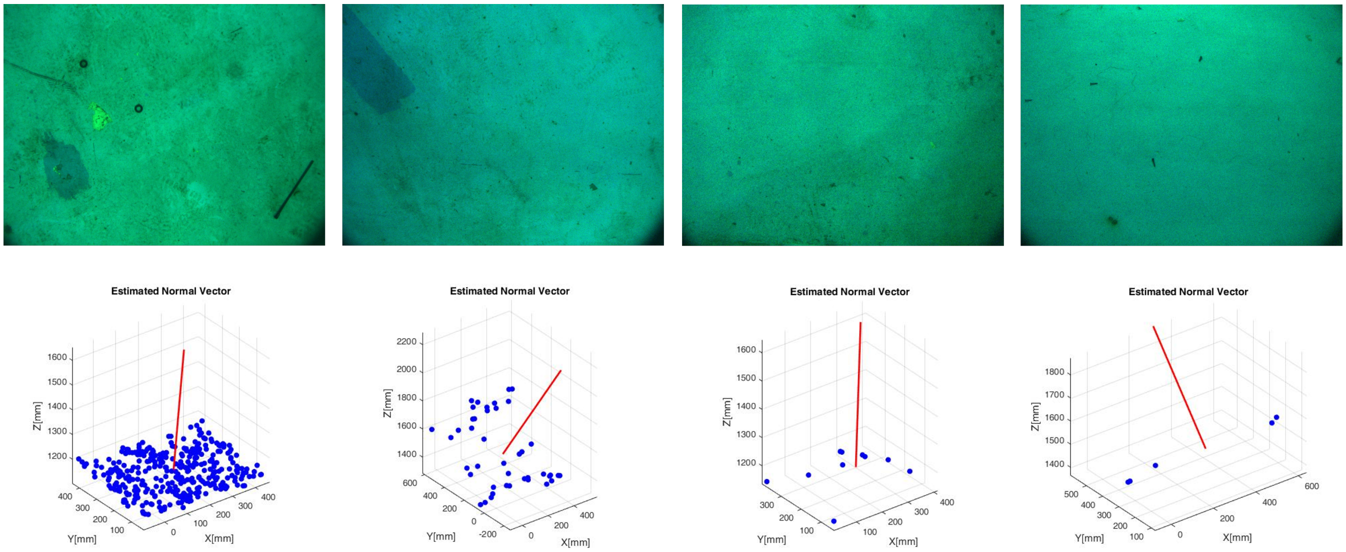
$$\begin{aligned} d &= \frac{|d|}{\sqrt{a^2 + b^2 + c^2}} \\ \phi &= atan2(\frac{a}{c}) \\ \theta &= atan2(\frac{sgn(c) * b}{\sqrt{a^2 + c^2}})5 \end{aligned} \tag{6}$$

Then, the camera heading can be aligned to the surface normal direction by controlling the vehicle motion using the distance and yaw angle to the plane, and also controlling the camera tilt unit using the pitch angle to the plane.

4. Experimental Test

4.1. Experimental Setup

An experiment was conducted in an indoor water tank in KAIST shown in Fig. 6. An underwater stereo camera unit was made (See Fig. 7). The stereo camera unit consists of a stereo camera (Pointgrey Bumblebee2), its waterproof housing, and a revolute joint for tilt motion. For calibration and rectification of the stereo camera, check board images were taken and the intrinsic and extrinsic parameters were acquired by using calibration toolbox in [7]. To imitate the underwater hull surface scene, the stereo camera faced down to the floor of the water tank from the distance of 1m. The images were taken in various places from feature-rich regions where there are many scratches and strains to feature-poor regions where the floor is clean. To check if the algorithm can reliably estimate the relative orientation to the surface, images were taken in various angles by rotating the revolute joint in camera module.



(a) (b)

Fig. 8: Images from the experimental data and their corresponding normal vector estimation results in a feature-rich region (a) and in a feature-poor region (b).

TABLE 1: Experiment Result.

No. of matched features	Actual relative angle [deg]	Angle estimation error [deg]
1197	0	-3.41
1263	-45	-5.33
103	0	-0.90
120	-30	-5.85
27	-0	-0.94
26	-45	-9.76

4.2. Results

Figure 8 shows the estimated normal vector of the plane in a feature-rich environment (a) and in a feature-poor environment (b). Table 1 shows the error of the estimated angle with respect to the number of matched features and the ground truth angle of the floor relative to the camera. The error angles are less than 10 degree, however the experimental results show that the error increases when the angle between the camera heading and the surface normal gets larger. By looking closely to the matching result, the mismatched features are mostly placed far from the camera. This mismatching happens due to attenuation and scattering of visual light, as the target surface gets further due to tilted angle. Also, the locations of features lying on left image differs to the right image due to light conditions and image transformations which requires the consideration of uncertainty of feature location.

5. Conclusion

In this study, we constructed a stereo vision system for underwater pose estimation relative to the hull surface which is crucial for autonomous underwater hull inspection. The proposed algorithm used visual salient features detected and matched by SURF algorithm to find stereo correspondence, RANSAC with the plane model to reject outliers and SVD to estimate the plane equation. For validation, an experiment was carried out in a water tank. The effectiveness of the proposed stereo vision system was verified through the various scenes from feature-poor scenes to feature-rich scenes.

Acknowledgement

This research was a part of the project titled 'Development of an autonomous ship-hull inspection system', funded by the Ministry of Oceans and Fisheries, Korea.

References

[1] S. Hong, J. Kim, J. Pyo and S. Yu, "A robust loop-closure method for visual SLAM in unstructured seafloor environments," Autonomous Robots, Vol. 40, No. 6, pp. 1095-1109, 2016.

[2] A. Elibol, J. Kim, N, Gracias and R, Garcia, "Efficient image mosaicing for multi-robot visual underwater mapping," Pattern Recognition Letters, Vol. 46, No. 1, pp. 20-26, 2014.

[3] P. J. Vaganay, M.L. Elkins and S. Willcox, Ship hull inspection by hull-relative navigation and control, In proceedings of OCEANS 2005 MTS/IEEE, Vol. 1, pp. 761-766, 2005.

[4] C. Harris and M. Stephens, A combined corner and edge detector, In Alvey vision conference, Vol. 15, Manchester, UK, 1988.

[5] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.

[6] H. Bay, T. Tuterlaars and L. V. Gool, SURF: Speed Up Robust Features, Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-349, 2008.

[7] J. Y. Bouguet, Camera calibration toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/.